Maximum Likelihood Estimation

1/27/22

Binary R.V.: $X_i = \begin{cases} 0 & \text{"failure"} \\ 1 & \text{"success"} \end{cases}$ $i = 1, 2, \dots, n$

- Simple random sample from large population of n people:

$X_i = \begin{cases} 1 & \text{agree} \\ 0 & \text{disagree/neutral} \end{cases}$ ← answer of i^{th} person

$P(X_i = 1) = \pi$ ⟶ true proportion in population that would agree if asked

Summary statistic: $Y = X_1 + \dots + X_n \sim Bin(n, \pi)$

Observe $Y = y$ + want to estimate π.

Prior to data collection → Y was random

$P(Y = y) = \binom{n}{y} \pi^y (1-\pi)^{n-y}$ $y = 0, 1, \dots, n$

↳ $f(y \mid \pi)$

Post-data collection → Know value of y; want to estimate π

$L(\pi \mid y) = \binom{n}{y} \pi^y (1-\pi)^{n-y}$ $0 < \pi < 1$

Maximum likelihood → choose the value of π that maximizes the probability of observing y for given π

Almost always easier to maximize log-likelihood:

$$\ell(\pi \mid y) = \log L(\underline{\pi} \mid \underline{y}) = \log \binom{n}{y} + y \log \pi + (n-y) \log(1-\pi)$$

maximize wrt this

fixed, known

Differentiate: $\frac{\partial}{\partial \pi} \ell(\pi \mid y) = 0 + y\left(\frac{1}{\pi}\right) + (n-y)\left(\frac{1}{1-\pi}\right)(-1) \overset{\text{set}}{=} 0$

$$\Rightarrow \hat{\pi} = \frac{y}{n}$$

Example: $X = \begin{cases} 1 & \text{made a NY resolution in 2022} \\ 0 & \text{not} \end{cases}$

$n = 14$ $\quad x = 5$ $\quad \pi$ = true prob. of making a NY resolution

$$L(\pi \mid x=5) = \binom{14}{5} \pi^5 (1-\pi)^{14-5} \qquad 0<\pi<1$$

Exact Binomial Inference : Model: $Y \sim Bin(n, \pi)$

Common for inference about π:

Large n / SD

Statistic: $\hat{\pi} = \frac{Y}{n} \overset{\cdot}{\sim} N\left(\pi, \sqrt{\frac{\pi(1-\pi)}{n}}\right)$

p-value = prob. of observed data, or more extreme, assuming $\pi = \pi_0$.

Ex (cont.): $H_0: \pi = 0.40$ $\quad$ Observe: $y = 5$

$H_a: \pi < 0.40$ $\quad n = 14$

p-value: $P(Y \leq 5 \mid \pi = 0.40)$

Exact $\quad$ Approx

R: pbinom(5, 14, 0.40) $\quad$ pnorm($5/14$, 0.40, $\sqrt{\frac{0.4(0.6)}{14}}$)

Chapter 2: Contingency Tables (Two-way tables)

(Skip 1.4–1.5)

Y — $I \times J$ table

X \ Y	1	2	3	...	J	Total
1	n_{11}	n_{12}	n_{13}	...	n_{1J}	n_{1+}
2	n_{21}	n_{22}		...		n_{2+}
$\vdots$	$\vdots$					$\vdots$
I	n_{I1}				n_{IJ}	
Total	n_{+1}	n_{+2}		...	n_{+J}	n

n_{ij} = count (frequency) in i^{th} row & j^{th} col.

n_{i+} = row i total $\qquad$ n_{+j} = col. j total

Probabilities: Joint probability: $\pi_{ij} = P(X=i, Y=j)$

prob. of ending up in $(i,j)^{th}$ cell

Marginal probability:

$$\pi_{i+} = P(X=i) \qquad \pi_{+j} = P(Y=j)$$

Conditional probability: (no notation in book)

$$\pi_{i|j} = P(X=i \mid Y=j)$$

or $$\pi_{j|i} = P(Y=j \mid X=i)$$

Estimates?

$$\hat{\pi}_{ij} = \frac{n_{ij}}{n}$$

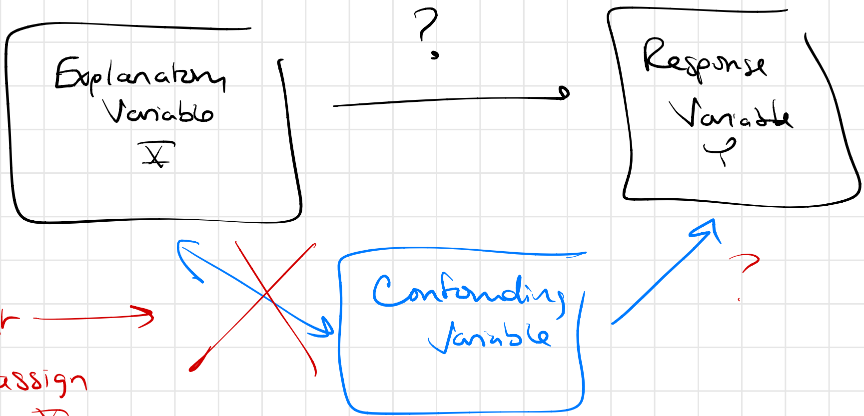
$$\hat{\pi}_{i+} = \frac{n_{i+}}{n} \qquad \hat{\pi}_{+j} = \frac{n_{+j}}{n}$$

Conditional:

$$\hat{\pi}_{j|i} = \hat{P}(Y=j \mid \underbrace{X=i}_{\text{subset}}) = \frac{n_{ij}}{n_{i+}}$$

subset — restricted to i^{th} row

Confounding variables:



experiment →

randomly assign values of X